AMENDMENTS TO THE SPECIFICATION

Please revise paragraphs 0032 and 0055 of the specification as follows:

[0032] As illustrated in Figure 1, the prefetch ~~client 58~~ client 56 may access a prefetch cache 58, which may be implemented with random access memory and/or hard disk storage of the web server machine(s) 32. The prefetch cache 58 is preferably used to cache data retrieved from the prefetch service 50 in order to reduce the frequency with which queries are sent to the prefetch service. For example, the cache 58 may store the subtask profile data for some or all of the most frequently accessed web pages, so that requests for these pages can be processed on most page requests without querying the prefetch service 50. Each entry in the prefetch cache 58 may correspond to a particular entry in the task-to-subtask mapping table 52, and may be created or updated based on data returned by the prefetch service. In one embodiment, the prefetch service 50 assigns a time-to-live (TTL) value to each such cache entry to specify how long the associated entry may be used. The TTL value for a given task ID may be generated based on collected historical data indicative, for example, of how often new subtasks are added to the associated subtask profile 52B.

[0055] If the prefetch service 50 is replicated across multiple ~~servers 52B~~ servers 50B, each such server can update its own respective task-to-subtask mapping table 52 based on the feedback it receives. Because each such ~~server 52B~~ server 50B will receive similar feedback data over time for like tasks, these tables will evolve similarly over time, and will thus contain similar data at any given point of time. As a result, each instance of the prefetch service 50 can operate independently of other instances. For web pages that are accessed infrequently (e.g., once per hour), the prefetch service 50 may instruct the prefetch client 56 to send the feedback data to all of the prefetch servers 50B, such that each such server may update its respective table 52.